MOLECULAR CELL BIOLOGY

Revised Printing with Expanded Index

JAMES DARNELL

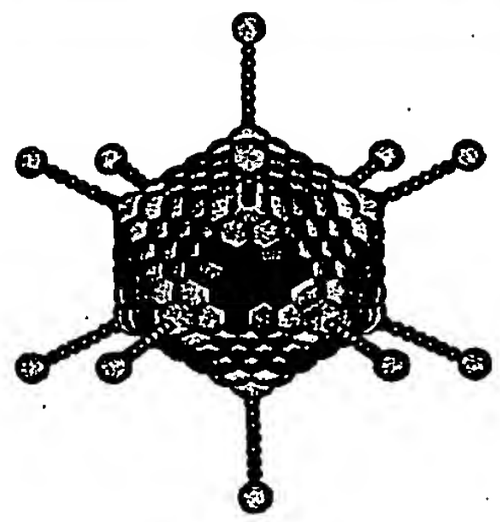
Vincent Astor Professor
Rockefeller University

HARVEY LODISH

Member of the Whitehead Institute for Biomedical Research
Professor of Biology, Massachusetts Institute of Technology

DAVID BALTIMORE

Director of the Whitehead Institute for Biomedical Research
Professor of Biology, Massachusetts Institute of Technology



SCIENTIFIC AMERICAN BOOKS

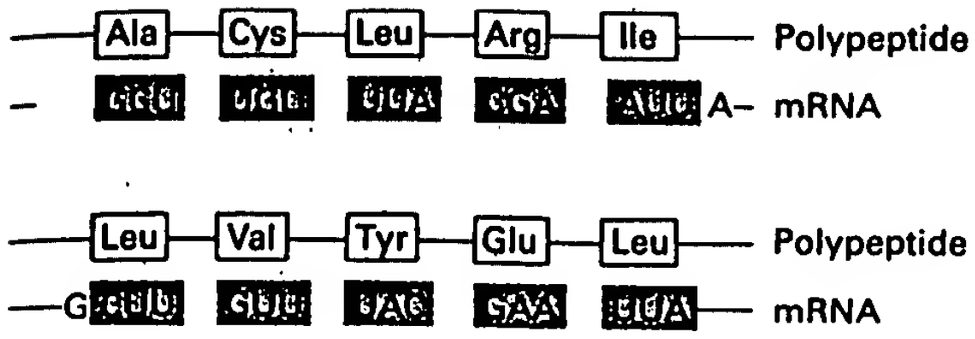


Figure 4-5 An overlapping triplet code that is read in two different frames. The mRNA is the same sequence in both lines but is read in a different "frame." Such codes have been discovered in the DNA of viruses that infect bacterial and mammalian cells.

The different codons for a given amino acid are said to be *synonymous*. The code itself is termed *degenerate*, which means simply that it contains redundancies. However, this degeneracy does not cause ambiguity in translation, because each triplet codes for only one amino acid.

The "start" *(initiator)* codon AUG specifies the amino acid methionine: all protein chains in prokaryotic and eukaryotic cells begin with this amino acid. At the beginning of a few chains, methionine is encoded instead by GUG. The three codons UAA, UAG, and UGA do not specify amino acids but constitute "stop" *(termination)* signals at the ends of protein chains. So a precise linear array of ribonucleotides grouped into threes in the mRNA specifies a precise linear sequence of amino acids in a protein, and also signals to ribosomes where to start and stop synthesis of the protein chain.

Table 4-1 The genetic code*

First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop (och)	Stop	A
	Leu	Ser	Stop (amb)	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Met)	Ala	Glu	Gly	G

*Bases are given as ribonucleotides, so U appears in the table instead of T. "Stop (och)" stands for the ochre termination triplet, and "Stop (amb)" for the amber. AUG is the most common initiator codon; GUG usually codes for valine, but it can also code for methionine to initiate an mRNA chain.

Synthetic mRNA and Trinucleotides Break the Code

The discovery of mRNA and how it functions led to the solution of the genetic code, one of the great triumphs of modern biochemistry. The underlying experimental work on mRNA and the code was largely carried out with the use of cell-free extracts from bacteria. All the necessary components for protein synthesis except mRNA (tRNAs, ribosomes, amino acids, and energy-rich nucleotides—ATP and GTP) were present in these extracts. Upon the addition of chemically defined synthetic mRNAs, the extracts formed specific polypeptides. For example, synthetic mRNA composed only of U residues yielded polypeptides made only of phenylalanine. Thus, it was concluded that UUU codes for phenylalanine. Each of the other three homopolymers likewise coded for a single amino acid (Figure 4-6). Next, synthetic mRNA that has alternating bases was used; for example,

. . . A C A C A C A C A C A C A . . .

The polypeptides made in response to this polymer contained alternating threonine and histidine residues. But this result alone was not enough to determine whether threonine was encoded by ACA and histidine by CAC, or

Table 4-2 The degeneracy of the genetic code

Number of synonymous codons	Amino acid	Total number of codons
6	Leu, Ser, Arg	18
4	Gly, Pro, Ala, Val, Thr	20
3	Ile	3
2	Phe, Tyr, Cys, His, Gln, Glu, Asn, Asp, Lys	18
1	Met, Trp	2
Total number of codons for amino acids		61
Number of codons for termination		3
Total number of codons in genetic code		64

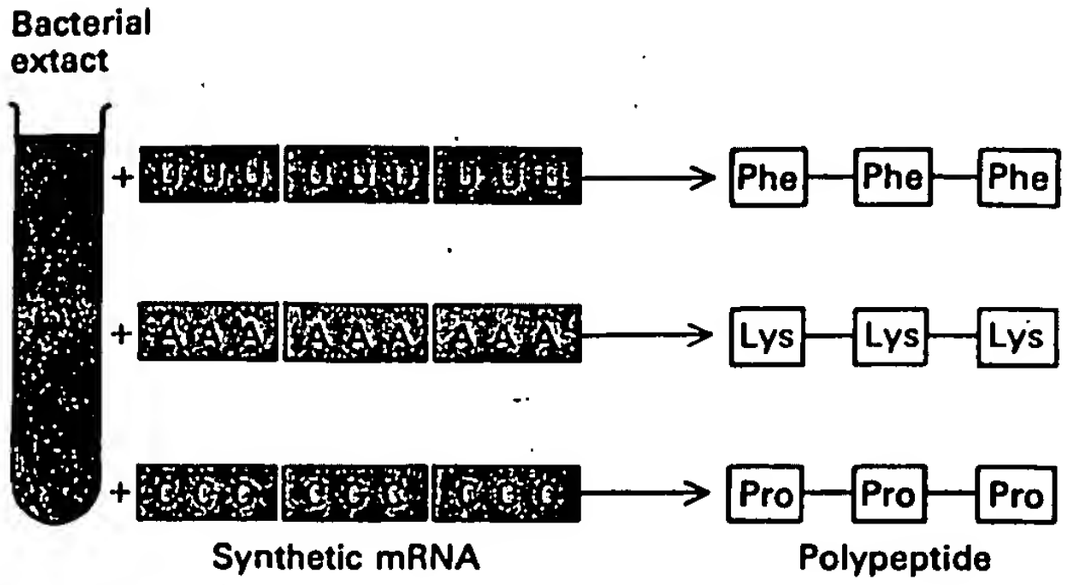


Figure 4-6 The genetic code was worked out largely by using bacterial extracts that contained all the components necessary for protein synthesis except mRNA. When synthetic mRNAs consisting entirely of a single type of nucleotide were added to the extracts, polypeptides composed of a single type of amino acid were formed. For example, polyphenylalanine was formed when polyuridylic acid was added. (Polyguanylic acid, G—G—G— . . . , encodes polyglycine but does so very poorly, because of the tendency of poly G to "stack" or undergo intrachain interactions.) [See M. W. Nirenberg and J. H. Matthei, 1961, *Proc. Nat'l Acad. Sci. USA* 47:1588.]

vice versa. A further experiment was necessary. An mRNA made of repeated sequences of AAC,

. . . A A C A A C A A C A A C A . . .

stimulated the synthesis of three kinds of polypeptide chains: all asparagine, all threonine, and all glutamine. Apparently, the decoding mechanism could start at any nucleotide, so that it could read the mRNA as three different repeated codons: all AAC, all ACA, or all CAA. The only codon in common between the two-codon mRNA and the three-codon mRNA was ACA, and the only amino acid in common in the polypeptide products was threonine. Therefore, ACA was assigned to threonine (Figure 4-7). Comparisons of the coding capacity of many such mixed polynucleotides revealed a substantial part of the genetic code.

In addition to these experiments, which used synthetic nucleic acids that instructed bacterial extracts to synthesize specific polypeptides, another type of experiment with extracts also was of great importance in the solution of the code. This experiment (Figure 4-8) showed specific trinucleotides would bind individual tRNAs to ribosomes, allowing codon assignment to each trinucleotide.

In all of the experiments in which bacterial extracts were programmed with synthetic mRNAs, the rate of formation of polypeptides was much lower than the rate when natural mRNAs were added to bacterial extracts. This was because the synthetic mRNA lacked start codons and stop codons. It was later appreciated that whereas the coding ability of the synthetic mRNA produced reliable results in experiments designed for the purpose of deciphering the code, it was only when natural mRNAs were added to the bacterial extracts that true proteins were programmed by mRNA. The first successful synthesis of a specific protein occurred when the mRNA of bacteriophage F2 was added to bacterial extracts, and the coat, or capsid, protein (the "packaging" protein that covers the virus particle) was formed.

Transfer RNA Decodes mRNA by Base-Pairing Codon-Anticodon Interactions

There is no evidence for direct chemical recognition between specific nucleic acid bases and specific amino acids. That is, in the synthesis of a polypeptide chain, the trip-

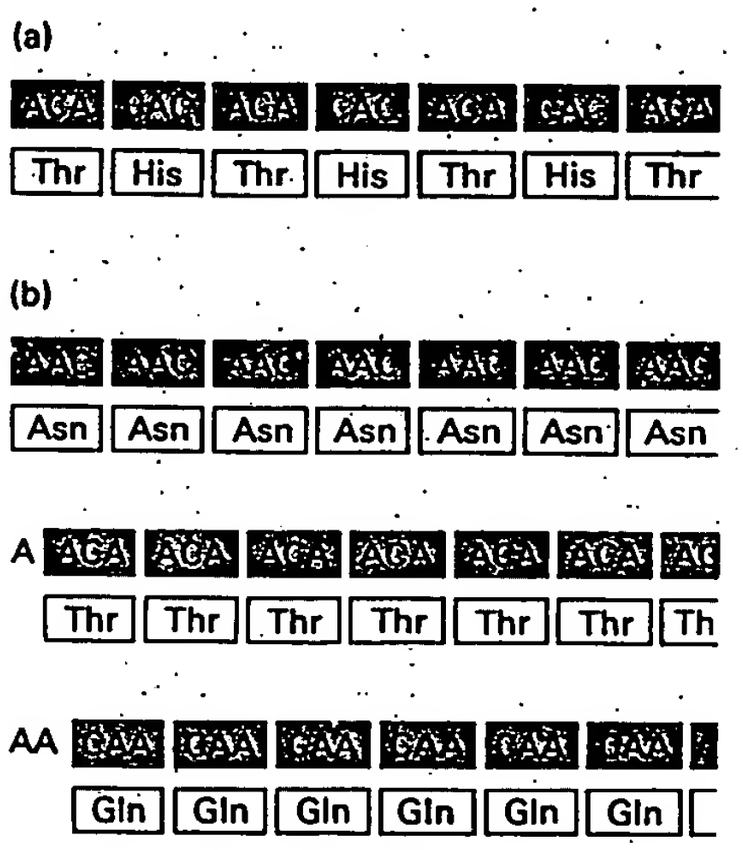


Figure 4-7 When synthetic mRNAs made with alternating A and C residues were added to a bacterial extract, polypeptides made of alternating threonine and histidine residues were formed. The assignment of threonine to ACA and histidine to CAC, as shown in (a), was made possible by another experiment: Synthetic mRNAs made with repeating sequences of

. . . AACAAC . . .

yielded three different kinds of polypeptide chains—all asparagine, all threonine, and all glutamine. Because the only codon in common in the two experiments was ACA, and the only amino acid product in common was threonine, threonine could be assigned to ACA. The other two assignments in (b) (asparagine to AAC and glutamine to CAA) were derived from further experiments. The "cracking" of the genetic code was thus a laborious, step-by-step process. [See H. G. Khorana, 1968, in *Nobel Lectures: Physiology or Medicine (1963–1970)*, Elsevier (1973), p. 341.]